A Temporal Kernel Method to Compute Effective Radiative Forcing in CMIP5 Transient Simulations*

Erik J. L. Larson

CIRES, University of Colorado Boulder, and Chemical Sciences Division, NOAA/ESRL, Boulder, Colorado

Robert W. Portmann

Chemical Sciences Division, NOAA/ESRL, Boulder, Colorado

ABSTRACT

Effective radiative forcing (ERF) is calculated as the flux change at the top of the atmosphere after allowing rapid adjustments resulting from a forcing agent, such as greenhouse gases. Rapid adjustments include changes to atmospheric temperature, water vapor, and clouds. Accurate estimates of ERF are necessary in order to understand the drivers of climate change. This work presents a new method of calculating ERF using a kernel derived from the time series of a model variable (e.g., global mean surface temperature) in a model-step change experiment. The top-of-atmosphere (TOA) radiative imbalance has the best noise tolerance for retrieving the ERF of the model variables tested. This temporal kernel method is compared with an energy balance method, which equates ERF to the TOA radiative imbalance plus the scaled surface temperature change. Sensitivities and biases of these methods are quantified using output from phase 5 of the the Coupled Model Intercomparison Project (CMIP5). The temporal kernel method is likely more accurate for models in which a linear fit is a poor approximation for the relationship between temperature change and TOA imbalance. The difference between these methods is most apparent in forcing estimates for the representative concentration pathway 8.5 (RCP8.5) scenario. The CMIP5 multimodel mean ERF calculated for large volcanic eruptions is 80% of the adjusted forcing reported by the IPCC Fifth Assessment Report (AR5). This suggests that about 5% more energy has come into the earth system since 1870 than suggested by the IPCC AR5.

1. Introduction

Accurate radiative forcing estimates are necessary to understand the global energy budget and future climate responses to anthropogenic and natural emissions. Estimating the radiative forcing that is driving climate change is difficult because of the complex climate response to the many individual forcing agents. It is currently impossible to directly measure the globally averaged radiative forcing. Observations from space measure the top-of-atmosphere (TOA) radiative imbalance, which includes both the forcing and Earth's thermal response. Models are used to calculate the radiative forcing, with several different definitions of forcing based on the modeling method employed.

Chapters 7 and 8 of the IPCC Fifth Assessment Report (AR5) (Boucher et al. 2013; Myhre et al. 2013) include a thorough discussion of the different radiative forcing (RF) definitions and the methods used to calculate them. Several methods to compute RF have been proposed, which utilize models of varying complexity. The instantaneous and adjusted forcing has traditionally been computed offline with radiation codes using a model or climatological atmosphere and the forcing perturbation. The instantaneous forcing keeps all fields constant (except the forcing agent), while other forcing definitions allow for the atmosphere to adjust. The "adjusted forcing" allows only stratospheric temperature adjustments. While informative, these definitions make it difficult to estimate RF of complicated climate

processes such as aerosol indirect effects. The fixed sea surface temperature (SST) technique was developed to allow these complex responses to be considered in the forcing estimate (Shine 2003; Hansen et al. 2002). In addition, this method allows more rapid adjustments to respond to the forcing. Rapid adjustments respond before global mean temperatures change and include changes to atmospheric temperature, water vapor, and clouds (IPCC 2013). This method calculates the forcing by differencing two climate model results with fixed sea surface temperatures after they have run to equilibrium. The forcing agent is the only change between the model runs. None of these techniques, however, allow the forcing to be calculated in a transient climate simulation. This makes it difficult to compute the RF differences between model runs and limits the analysis that can be carried out (e.g., estimating Earth's energy balance in models).

Gregory et al. (2004) developed a method to calculate the RF and climate feedback parameter that relies on a linear forcing–response relationship. It was noted that the relationship between the TOA radiative imbalance N and global mean surface atmospheric temperature change ΔT in a step-change CO_2 simulation was approximately linear,

$$N = F - \alpha \Delta T, \tag{1}$$

and the slope obtained by regressing N against ΔT gives an estimate of the climate feedback parameter α. The y intercept gives an estimate of the radiative forcing F. Hansen et al. (2005) suggested that a regression length of 10–30 yr was needed to retrieve a forcing that was consistent with the effective forcing in the fully coupled GISS ModelE (used in CMIP3). Andrews et al. (2012) applied this regression to the database from phase 5 of the Coupled Model Intercomparison Project (CMIP5) and found that 150-yr regressions retrieved an initial forcing that was consistent with the values calculated using fixed SST simulations.

Forster and Taylor (2006) applied the linear relationship in Eq. (1) to transient model values of N and ΔT to estimate a time series of RF. This assumes that the climate feedback parameter is constant during the transient calculation. Usually α is estimated from step change experiments using the same model. We refer to this method of calculating RF as the energy balance (EB) method below.

Numerous studies have shown that the assumption of a constant feedback parameter is reasonable for short time scales (~10–100 yr), whereas it breaks down on longer time scales (Andrews et al. 2012; Forster et al. 2013; Williams et al. 2008; Hansen et al. 2005; Gregory

et al. 2004; Senior and Mitchell 2000). This is due to climate feedbacks changing the relationship between the global mean temperature and TOA imbalance on long time scales. Thus, the retrieved feedback parameter and initial forcing are dependent on the length of time used in the regression.

We propose a new method of calculating the radiative forcing in transient simulations using a temporal kernel (TK) function derived from step change model calculations (e.g., $4 \times CO_2$ experiment). We are using the term temporal kernel to distinguish from other radiative kernel methods used to calculate feedbacks (Soden et al. 2008; Shell et al. 2008). We demonstrate that this TK method is consistent with the energy balance method for calculating forcing but does not have the same biases. In particular, it does not assume a constant climate feedback parameter.

There is not a consensus on the term for the radiative forcing calculated from transient simulations. Historically, adjusted forcing has referred to forcing in which the only adjustments were to stratospheric temperatures (IPCC 1990; Hansen et al. 1997). Others (Andrews et al. 2012; Forster et al. 2013) used the term adjusted forcing to describe the forcing calculated with the EB method, which allows for other rapid adjustments including water vapor and clouds. Given the consistency between the initial forcing of the EB method and fixed SST method of calculating forcing, we follow the example from the IPCC AR5 (IPCC 2013) and use the term effective radiative forcing (ERF) to indicate the forcing calculated in transient climate simulations by the EB method or the TK method described below.

In this paper we apply the EB method to the CMIP5 output and quantify the effects of using different length regressions. We also present a new temporal kernel method to calculate ERF and compare it with the EB method, assessing the strengths and weaknesses of each. Finally, we use these methods to calculate the forcing of volcanic eruptions and consider their impact on the global energy budget.

2. CMIP5 processing

To estimate the ERF in the CMIP5 models [described in Taylor et al. (2012)], we first need to remove drifts and offsets from the CMIP5 model output. Monthly output of the models' first realization (r1i1p1) is analyzed for each available CMIP5 model as of January 2015. Fluxes for the solar downwelling, solar upwelling, and longwave upwelling at the TOA (CMIP5 parameters rsdt, rsut, and rlut, respectively) are used to calculate the N. The CMIP5 models are often subject to drift (Forster et al. 2013; Gupta et al. 2013), which is necessary to
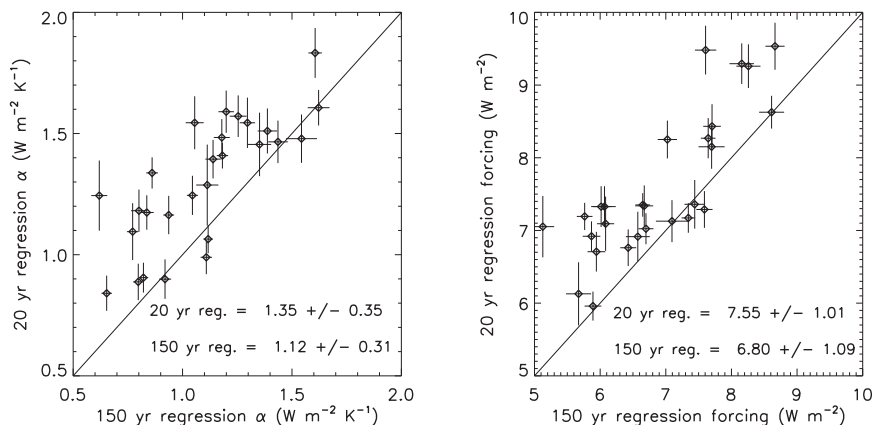
FIG. 1. Scatterplot of the (left) α and (right) F_0, retrieved with 20- and 150-yr regressions. Error bars represent 1σ uncertainties from the regressions. The solid line has a slope of 1.

remove. We use the preindustrial control (piControl) simulations to reduce model drifts using the same time period as the historical simulations. Most models specify the year the historical simulation is branched from the piControl simulation; otherwise the beginning of the piControl is used. There appears to be very little drift in N in the piControl simulations; however, there is substantial offset. We subtract the mean TOA imbalance in the first 10 years of the historical run from the historical and abrupt4×CO$_2$ simulations, thus making N an anomaly to the beginning of the historical simulation. We remove the linear trend in temperature in the piControl from the historical and abrupt4×CO$_2$ temperatures. The temperatures at the start of some historical simulations are offset from the piControl simulations, so we also subtract the mean from the first 10 years of the historical simulation, thus making temperatures anomalous to the start of the historical simulation. Similarly, we remove the linear drift from the piControl and the offset from the first 10 years of the historical simulation for the ocean heat content.

We have chosen to make the output variables [N, temperature (T), and ocean heat content (OHC)] anomalies compared to the beginning of the historical simulation as opposed to the preindustrial control. This was done intentionally because in many models there is a discrepancy between the beginning of the historical simulation and the preindustrial control. Some of these discrepancies are due to incorrect labeling of the branch year. The inconsistencies in temperature are up to a half of a degree, which is substantial especially when calculating integrated quantities. For consistency between models and to include the most models possible, we have made all output variables anomalous to the start of their historical simulation. Since not all of the models have

the same start date for their historical simulations, later in the paper we report the forcing relative to 1870–80.

3. Energy balance method

Following the method of Forster et al. (2013), we compute the forcing time series of the CMIP5 historical simulations using the EB method. We use the abrupt4×CO$_2$ simulation from each model to estimate the climate feedback parameter using Eq. (1). Because of the nonlinearity in the relationship between TOA radiative imbalance and global mean surface temperature change, the number of years used in the regressions has a substantial (up to 20%) effect on the retrieved climate feedback parameter and thus the radiative forcing. We compare climate feedback parameters and initial forcing using regressions of 20 and 150 yr in Fig. 1. The length of regression that is most consistent with the fixed SST forcing is model dependent. Of the seven models with fixed SST ERFs calculated by Andrews et al. (2012), three are more consistent with a 20-yr regression, one is more consistent with a 150-yr regression, and three are between the two regressions. We find that longer regressions are influenced by the nonlinear relationship between temperature and TOA radiative imbalance and estimate a smaller climate feedback parameter and initial ERF. Shorter regressions estimate larger climate feedback parameters and initial ERF and have larger uncertainty. The multimodel mean feedback parameter and standard deviation is 1.35 ± 0.35, 1.21 ± 0.32, and 1.12 ± 0.31 W m^{-2} K^{-1} for the 20-, 100-, and 150-year regressions, respectively. The error bars indicate the 1-sigma uncertainty of the regression. Our 100-yr regression mean feedback parameter is closest to the value adopted by Murphy et al. (2009), 1.25 ± 0.5 W m^{-2} K^{-1}, based on annual and interannual regressions of ERBE

and CERES data. However, their value is consistent with all three estimates from the CMIP5 multimodel mean. While most of the abrupt4×CO$_2$ experiments are run for only 150 years, the few that run for longer indicate a continued flattening of the relationship of N and ΔT and thus even smaller retrieved feedback parameters and initial forcings with longer regressions.

This section replicates some of the work in Andrews et al. (2012) and Forster et al. (2013). The main difference between our analysis and theirs is that we test different regression lengths on the retrieved parameters. The initial forcing and feedback parameters we calculate using the 150-yr regressions are generally within a few percent of those calculated by Andrews et al. (2012) and Forster et al. (2013) (see Table S1 in the supplemental material). The slight differences are due to the slightly different processing method we used to remove model drift and offsets. These different processing methods result in differences greater than 10% in estimated parameters in only two models, both of which have large differences between the temperature in the piControl simulation and the temperature at the start of the historical simulation. Our multimodel-mean feedback parameter and initial forcing is within 1% of that calculated by Forster et al. (2013).

4. Temporal kernel method of estimating forcing

We consider an alternative method to estimate the forcing that is based on a temporal kernel method. Good et al. (2011, 2013) created a simple model to calculate climate properties (surface temperature, precipitation, ocean heat uptake, etc.) given the time series of the forcing and the values of those properties from a step change experiment (e.g., abrupt4×CO$_2$). In their model, the climate property y in year i is equal to the convolution of responses to all previous annual forcing. Good et al. (2011, 2013) writes this finite convolution as the first of the following sums:

$$y_i = \sum_{j=0}^{i} x_j \frac{F_{i-j}}{F_0} = \sum_{j=0}^{i} x_{i-j} \frac{F_j}{F_0}, \quad (2)$$

where we have rewritten the summation in the second version. Here, y_i is the climate variable to be calculated in year i, x_{i-j} is the same climate property in the control simulation (abrupt4×CO$_2$), F_j is the forcing in year j, and F_0 is the initial forcing in the control (step change) experiment. We have chosen F_0 to be the initial forcing retrieved using regressions from the EB method. Thus, it is also dependent on the number of years used in the regression (see Fig. 1). We do this so that the forcing magnitudes are equivalent and the two methods can be

easily compared. However, one could use the forcing estimates from fixed SST calculations or the TOA imbalance in the first year of the abrupt4×CO$_2$ simulation.

Good et al. (2013) applied this equation as a computationally efficient way of exploring the response to forcing scenarios that have not been run by GCMs. This model is unique in that it can retrieve a range of variables based on their response in a control simulation and a given forcing time series. Good et al. also show that the method generally works better for globally averaged temperature than for precipitation.

We are interested in using the above relationship and equation to calculate the radiative forcing in an experimental run given the variable time series in the experimental and control run. To do this we note that Eq. (2) can be written as a matrix equation

$$\mathbf{Y} = \mathbf{X}\mathbf{F}/F_0, \quad (3)$$

where $\mathbf{Y}$ and $\mathbf{F}$ are vectors and F_0 is the forcing in the CO$_2$ step-change experiment. The term $\mathbf{X}$ is a matrix with elements defined as follows:

$$X_{i,j} = \begin{cases} 0, & \text{for } j > i \\ x_{i-j}, & \text{for } j \leq i \end{cases}. \quad (4)$$

Thus, by solving the linear system of equations or inverting the matrix $\mathbf{X}$ we can retrieve the following forcing:

$$\mathbf{F} = F_0 \mathbf{X}^{-1} \mathbf{Y}. \quad (5)$$

One issue with using a matrix inverse is that it can induce noise in the retrieved forcing. The amount of noise is dependent on the choice of variable used to construct matrix $\mathbf{X}$. We explore using three control variables to retrieve the forcing: temperature change ΔT, TOA radiative imbalance N, and ocean heat content change ΔOHC. To reduce the noise in the retrieval we fit a smooth function to the step change response for each variable. This fit is then used to construct the rows of the matrix $\mathbf{X}$ above. Another benefit of fitting the variable is that we can extend the time series beyond the length of the control experiment by extrapolating the fit. This was not necessary for the historical simulations but was for many models for representative concentration pathway 8.5 (RCP8.5) simulations. The variable N is well fit using a double exponential of the form

$$N = Ae^{-t/\tau_1} + Be^{-t/\tau_2}, \quad (6)$$

while ΔT and ΔOHC are well fit using the form

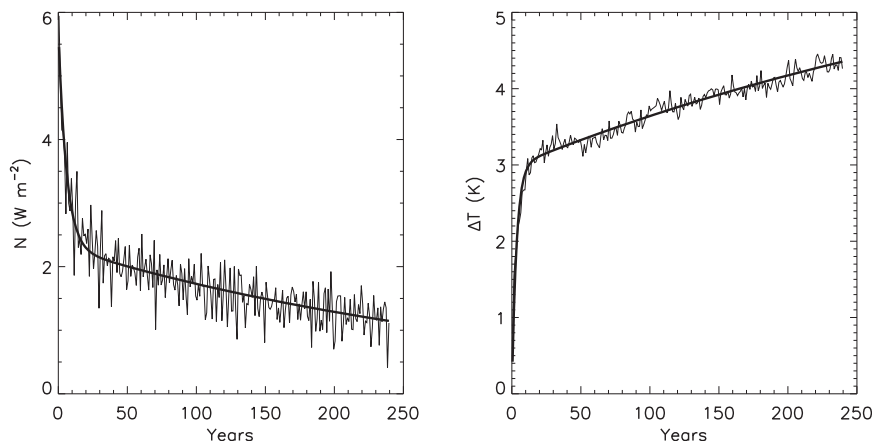$$\Delta T \text{ or } \Delta\text{OHC} = A(1 - e^{-t/\tau_1}) + B(1 - e^{-t/\tau_2}), \quad (7)$$

FIG. 2. (left) The TOA imbalance and (right) temperature change time series in the GFDL-ESM2M abrupt4×CO$_2$ simulation fit with the double exponential in Eqs. (6) and (7), respectively.

where A and B are amplitudes, t is time, and τ_1 and τ_2 are time coefficients representing the different response times of the variable to the forcing. Examples of the N and ΔT in the abrupt4×CO$_2$ GFDL-ESM2M simulation and their fits are shown in Fig. 2. The best fit for N uses values for A, τ_1, B, and τ_2 of 2.32, 340, 3.42, and 5.8, respectively. The best-fit coefficients will be different for every model. The mean values for A, τ_1, B, and τ_2 of the CMIP5 models are 2.6, 274, 3.5, and 5.8 for N.

We explore the ability of the TK method to retrieve the forcing using different control variables in Fig. 3 with the following approach. First, we apply the forward kernel method (Good et al. 2013) to the GHG forcing from Meinshausen et al. (2011) (red lines in Fig. 3, right) to retrieve the ΔT, N, and ΔOHC time series (solid black lines in Fig. 3, left). We then added autocorrelated noise, first-order autoregressive [AR(1)] model with autocorrelation 0.7 and amplitude 0.012 multiplied by the maximum of the variable time series, to the variable time series (dashed lines in Fig. 3, left). This simulates natural variability in these climate variables. We then apply our TK method to the time series of each variable (with added noise) to retrieve the forcing. The kernel used in Fig. 3 is calculated using parameters from the GFDL-ESM2M abrupt4×CO$_2$ simulation. The recovered GHG forcing time series (black lines in Fig. 3, right) is plotted with the original GHG forcing from Meinshausen et al. (2011). Note that the retrieved forcing would be equal to the original forcing if not for the presence of the added noise. The known forcing is recovered most accurately when using N to populate the kernel used in Eq. (5). Thus, we will use the TOA imbalance as our control variable to retrieve forcing time series in transient experiments in the rest of the paper. The TOA imbalance likely returns the best forcing because of

the shape of the response. The TOA imbalance has a large fast response and small slow response that decreases with time (Fig. 2, left). Alternatively, surface temperature has an opposite response in which the response to a change in forcing grows with time (Fig. 2, right). This growth amplifies the noise when applying the inverse [Eq. (5)].

The TK method of calculating ERF is not biased by the nonlinearity of the relationship between N and ΔT that influences the EB method. To demonstrate this, we used the EB and TK methods to retrieve the forcing time series from the abrupt4×CO$_2$ simulation. This has the advantage that the forcing should be constant throughout this simulation. In Fig. 4, we plot ERF estimated from the CCSM4 abrupt4×CO$_2$ simulation calculated using different methods and overlay the linear trend lines. ERF estimated with the EB method using a 150-yr regression to estimate α is quite flat over the 150-yr time period of the control run except for the first 10 years, which deviate substantially from zero and are indicated by a box in the figure. This indicates that the first 10 years are poorly fit with a linear regression of 150 years. ERF calculated with the EB method using a 20-yr regression to estimate α fit the first few decades very well but soon deviate from the constant forcing. There is significant slope, 0.66 W m^{-2} century^{-1}, in the ERF calculated using a 20-yr regression. The TK method, however, retrieves nearly constant ERF over the entire period of the abrupt4×CO$_2$ simulation. Not all models have nonlinearity in their regressions, and thus the slope is much less for some models. The multimodel mean slopes for the 150- and 20-yr regression EB method forcing are 0.088 and 0.418 W m^{-2} century^{-1}, respectively. The same slopes for the TK method are -0.025 and -0.028 W m^{-2} century^{-1}. This demonstrates the superiority of the kernel method for this metric.
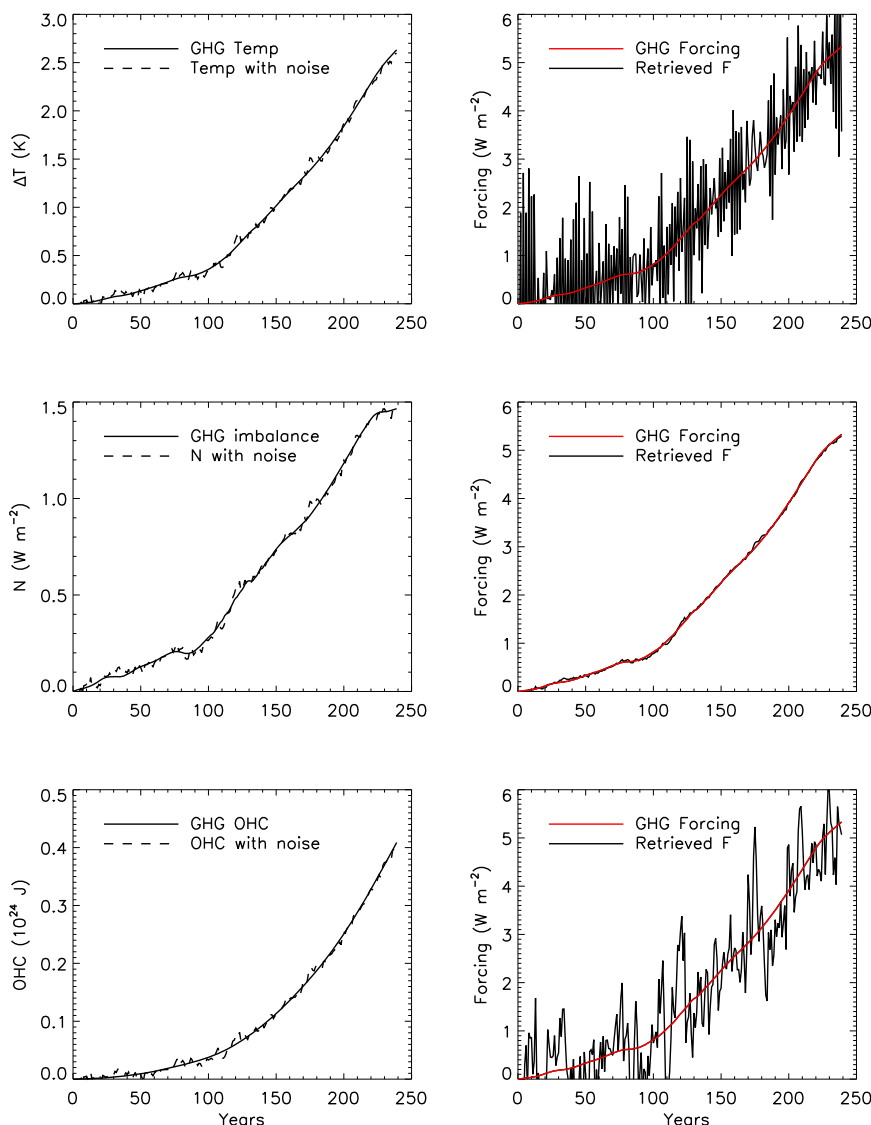
FIG. 3. The noise in the ERF calculated with the TK method is highly dependent on the variable used in the kernel. Of the three variables compared, (top) surface temperature, (middle) TOA imbalance, and (bottom) OHC, the TOA imbalance displayed the highest tolerance to noise. See text for details.

It should be noted that the TK method requires an estimate of the forcing in the control simulation, in this case the abrupt4×CO_2 simulation. This estimate can come from many different sources, including the imbalance from a fixed SST $4 \times CO_2$ simulation, the imbalance in the first year of the coupled $4 \times CO_2$ simulation, or the y intercept of the imbalance versus temperature regression. The ERF calculated from the TK method will scale linearly with this choice of F_0. We have chosen to use the y intercept from the imbalance versus temperature regression since this corresponds to the ERF calculated by the energy balance method. That way we can compare the two methods directly. The

forcing time series from the TK method in Fig. 4 is identical in the two plots except for the scaling.

5. Estimates of effective radiative forcing

One downside to the TK method is the larger noise or variability compared with the EB method. The TK method is sensitive to variation in N, which is larger than the variation in $N + \alpha \Delta T$ to which the EB method is sensitive. We compare the variability of these methods in Fig. 5. The multimodel means using 100-yr regressions to estimate α and F_0 are the solid (TK) and dashed (EB) lines, respectively. The difference in the means is small
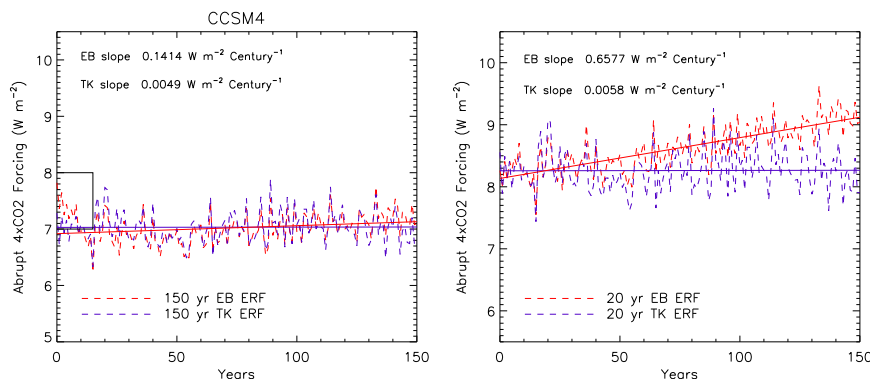
FIG. 4. Forcing from the CCSM4 abrupt4×CO$_2$ simulation retrieved with the TK and EB methods using parameters calculated from (left) 150- and (right) 20-yr regressions. A strong bias is present in the EB retrieved forcing using 20-yr regressions.

compared to the spread in the individual models. The light gray shaded region in Fig. 5 shows the 5th–95th percentile in the model-retrieved ERF calculated with the TK method for the 26 models that had output for the piControl, historical, and abrupt4×CO$_2$ simulations. The dash–dotted line represents the same percentiles for the EB method. The dark shading and dotted lines are 2 times the standard error of the means for the TK and EB methods, respectively. The standard error of the means is calculated from the spread of the means of 1000 iterations of Monte Carlo–sampled model ERF with replacement.

In Fig. 6 we show the multimodel mean ERF using the CMIP5 historical (and RCP8.5 after 2006 where available) simulations computed with the TK and EB methods. We also show estimates of ERF from Meinshausen et al. (2011) and Table AII.1.2 in AR5 (IPCC 2013). To directly compare the CMIP5 models with the IPCC AR5 and Meinshausen et al. we plot the forcing in Fig. 6 relative to 1870–80. Despite the biases in the EB method, the two methods generally agree to within a few percent in the multimodel mean retrieved ERF. The ERF from Meinshausen et al. generally agrees better with both of our retrieval methods than the IPCC estimates in AR5. This is especially true during volcanic events (see below for more discussion) but is also true during volcanically quiescent periods (e.g., 1950–62). This is mostly due to updated concentrations of forcing agents, especially anthropogenic and stratospheric aerosol, between the CMIP5 and publication of the IPCC AR5.

Between 1995 and 2015, a period of interest because it spans the "hiatus" in global warming (Allan et al. 2014; Smith et al. 2015; Trenberth and Fasullo 2013), the TK method retrieves an ERF that is on average 0.035 W m^{-2} larger than the EB method. As shown earlier, the choice of regression length will affect our estimates of ERF, with shorter regressions resulting in higher ERF

estimates both for the EB and TK method. The TK method is affected only because of the use of F_0 in Eq. (5). Between 1995 and 2015, the ERF calculated using parameters from 20-yr regressions is about 0.2 W m^{-2} larger than those calculated using parameters from 150-yr regressions (Fig. 6). This magnitude of forcing is comparable to changes in stratospheric water vapor or the background aerosol forcing that have been invoked to partially explain the hiatus period (Solomon et al. 2010, 2011).

The large bias in the EB method displayed in the CCSM4 output in Fig. 4 does not result in large differences in retrieved ERF in Figs. 5 and 6 for several reasons. First, not all models have this bias, so the multimodel mean reduces this effect. Second, assuming the bias is linear as in Fig. 4, then the magnitude of the bias in the EB method is proportional to the magnitude of ERF and the time since ERF. The ERF in Fig. 4 from the abrupt4×CO$_2$ experiment is much larger than the
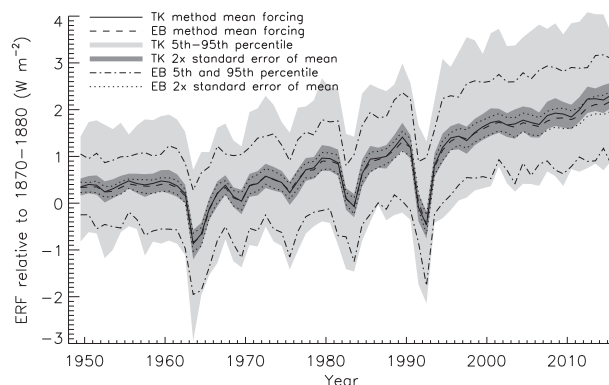

FIG. 5. Estimates of variability and uncertainty in the retrieved ERF from the two methods using parameters calculated from 100-yr regressions. The TK method ERF is more variable than the EB method ERF.
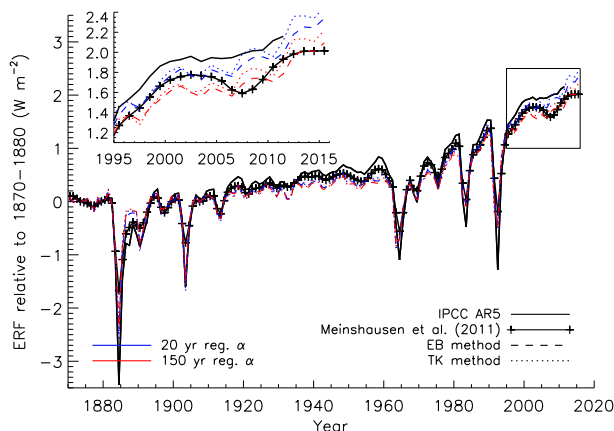
FIG. 6. ERF estimates from 1870 to 2015 retrieved from the TK and EB methods compared to Meinshausen et al. (2011) and IPCC AR5. The blue lines are calculated using parameters from 20-yr regressions, while the red lines are calculated using parameters from 150-yr regressions.

ERF in the historical simulation. Furthermore, in the historical simulation, ERF is dominated by recent changes, so the time for the bias to grow is short. Although we do not see a large difference between the forcing retrieval methods in the multimodel mean ERF in the historical simulation, we expect to see larger differences in individual models with larger nonlinearities, longer simulations, and simulations with larger forcing.

We can calculate the bias in the EB method by applying a linear model to the bias and using the abrupt4× CO_2 experiment to get the constant of proportionality. If the total ERF is simply the summation of the incremental ERF every year and the bias is proportional to the magnitude of the incremental ERF and the time since the ERF, then the total bias B can be written as follows:

$$B_i = A \sum_{j=0}^{i} (i - j)\Delta F_j. \qquad (8)$$

Here, B_i is the bias as a function of time step, ΔF_j is the incremental ERF in year j, and A is the amplitude. The value of A can be calculated from the abrupt4× CO_2 simulation as the slope in Fig. 4 divided by the magnitude of ERF.

Figure 7 shows RCP8.5 simulations that extend out to year 2300. There is much more divergence between the retrieval methods (red and blue lines) than found in the historical simulations. The bias calculated with Eq. (8) is shown as the dotted lines. The largest bias in the models presented is over 6 W m^{-2} in year 2300 from CSIRO Mk3.6.0. This model has a large nonlinearity in the regression between N and ΔT in the abrupt4× CO_2 simulation. Some models, such as the CNRM-CM5, have very little nonlinearity and the calculation methods

return similar ERFs. We expect the calculated bias added to the TK method ERF (bias plus TK is the dashed line) to be equal to the EB method ERF (red line). This expectation is upheld for the CSIRO Mk3.6.0 and MPI-ESM-LR models; however, in other models, the bias is overestimated. In some models (e.g., GISS-E2-H), the TK method actually retrieves a much larger ERF than the EB method, counter to our expectations. There are many possibilities for these discrepancies. One possibility is that the climatic response to a quadrupling of CO_2 might be different than the same amount of forcing from other sources or greenhouse gases. Also, the 4 × CO_2 simulations are generally run for 150 years, after which we have to extrapolate the imbalance to use the TK method. The climate may be in a different state after 300 years of CO_2 emissions and our extrapolation may not be correct. Similarly, the feedback parameter may change in unexpected ways over hundreds of years.

The difference in retrieved ERF between these two methods for different models indicates differing model responses to ERF. For instance, models that do not have large nonlinearity in their N and ΔT regressions likely respond similarly under different climate states. However, models that have large nonlinearities, and thus different ERF estimates with different methods, suggest that the feedback mechanisms are very different in a highly forced climate.

6. Volcanic conversion factor

The multimodel mean ERFs using both methods generally agree to within a percent of each other and to within 10% or better with that of Meinshausen et al. (2011) and the AR5 (IPCC 2013) outside of volcanic events (Fig. 6). However, volcanoes have substantially more negative forcing in the AR5 estimates than those of Meinshausen et al. and often our CMIP5 estimates. For example, we retrieve 0.9 W m^{-2} less (negative) forcing than the IPCC AR5 estimate in 1992 from the Mount Pinatubo eruption. To estimate the volcanic forcing, the IPCC AR5 scales the global mean volcanic stratospheric aerosol optical depth, originally from Sato et al. (1993) and since updated, by a factor of −25. This is the conversion factor corresponding to the adjusted radiative forcing (where only the stratospheric temperatures are allowed to adjust) calculated by the GISS ModelE (used in CMIP3) (Hansen et al. 2005). This is slightly higher than the conversion factor (−23) that the GISS ModelE needed to estimate ERF (Hansen et al. 2005). We call the ratio of these two factors (and forcing estimates) the rapid adjustment factor (RAF). For the GISS ModelE used in Hansen et al. (2005), they
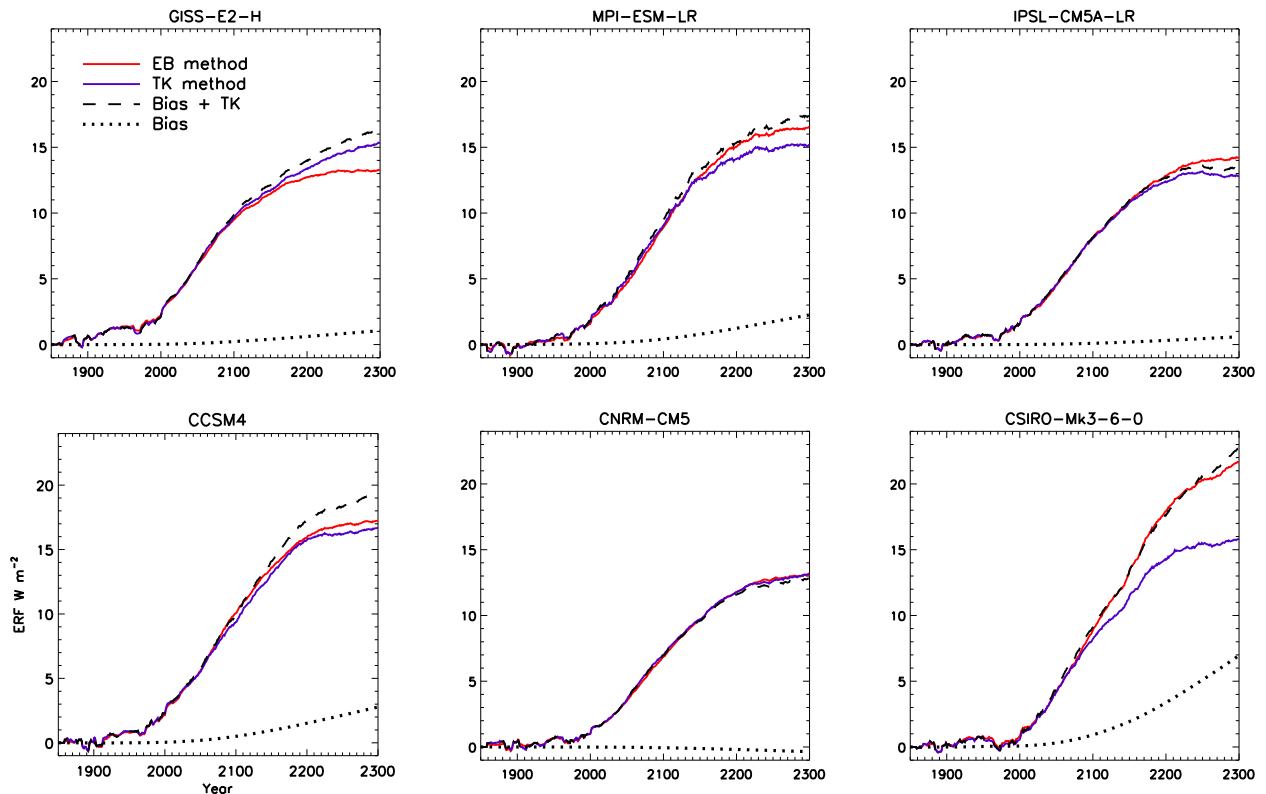
FIG. 7. ERF and bias [see Eq. (7)] in the historical and RCP8.5 simulations retrieved with the TK and EB methods using parameters derived from 20-yr regressions. See the supplementary material for plots using 150-yr regressions.

calculate a RAF of 0.92. This indicates that volcanoes have a positive rapid adjustment forcing in the GISS ModelE. Meinshausen et al. calculated the ERF by scaling the volcanic stratospheric aerosol optical depth by −23.5 and then multiplying by a factor of 0.7 to match the temperatures in their simple model to observed historical temperatures. The factor of 0.7 compensated for the excessive response due to volcanic aerosol in their simple model. Meinshausen et al. (2011) are essentially reporting ERF with a RAF for volcanic aerosols of 0.66, which explains why their values for volcanic forcing are substantially smaller than the IPCC AR5 estimates.

It is interesting to note that the CMIP5 model mean ERF using both calculation methods are similar to those of Meinshausen et al. (2011) for the Mount Pinatubo eruption. However, Fig. 6 shows that other volcanic events are more consistent with the IPCC AR5 forcing. We calculate the RAF in each model due to volcanic aerosol with the following method. We isolate each eruption by removing a line fit to 3 years before the eruption and 5–8 years after the eruption and total the remaining ERF. Then we use the ratio of the integrated ERF to the integrated volcanic aerosol optical depth from the GISS website (http://data.giss.nasa.gov/

modelforce/strataer/). This ratio represents the conversion factor needed between volcanic aerosol optical depth and effective radiative forcing for each model. By dividing our calculated conversion factors by the conversion factor for adjusted forcing (−25), we retrieve the RAF due volcanic aerosol for each model. Figure 8 shows the yearly averaged aerosol optical depth for the Mount Pinatubo eruption and the ERF divided by −25 for each model and retrieval method. The model RAFs for each method are reported in parentheses following the models. For the Mount Pinatubo eruption, we calculate a multimodel mean RAF of 0.7 (using 100-yr regressions to estimate α and F_0), which is substantially lower than the Hansen et al. (2005) value of 0.91 and more consistent with the value found by Meinshausen et al. (2011). Individual models may have substantial differences up to 0.35 in RAF as a result of noise in the calculated forcing and removal of the background trend. However, the multimodel mean RAF between these methods is surprisingly similar. Note that observed aerosol optical depth (AOD) peaks in 1992, while some models have ERF peaking in 1991, the year Mount Pinatubo erupted.

We repeated the analysis for the five largest volcanic events since 1880 and plot the retrieved RAF averaged over both methods for each model and volcano in Fig. 9,
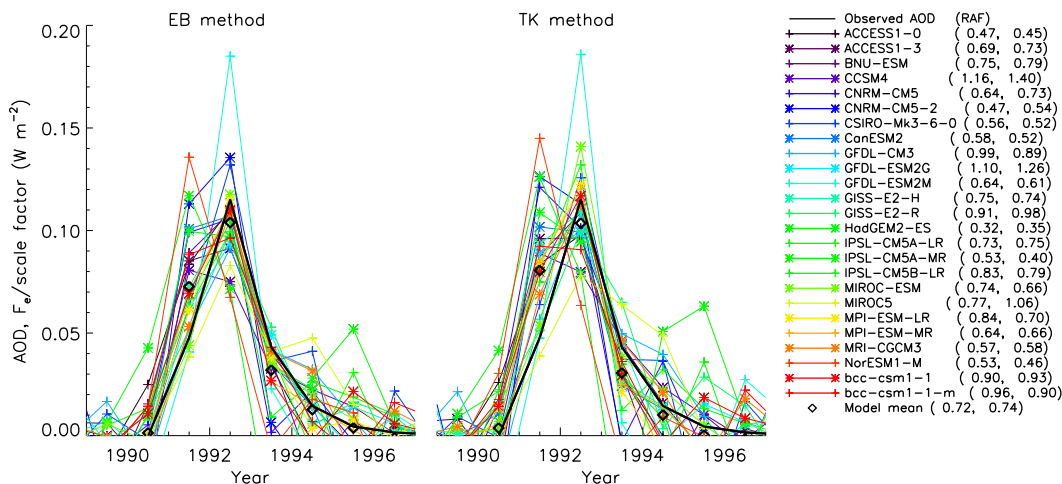
FIG. 8. Radiative forcing estimates of the Mount Pinatubo eruption isolated for each model. The integrated forcings are scaled to the observed AOD and used to retrieve the volcanic aerosol RAFs, which are shown in parentheses after the models. (Acronyms are available at http://www.ametsoc.org/PubsAcronymList).

along with the mean and standard deviation for each volcano. There is considerable scatter in the retrieved RAF for each volcanic eruption as a result of eruption size, location, and degree of temporal isolation of each eruption from the others. The different eruptions could have different climate responses and thus RAFs. However, because the standard deviations overlap in this analysis, the mean RAFs are statistically consistent between the eruptions. We have not found any evidence that the size or location of an eruption affects the RAF. We plot the models in order of increasing Krakatoa RAF along with trend lines. All trend lines slope up, indicating that models with a large Krakatoa RAF also have higher RAFs for other eruptions but with low r^2 values. The r^2 values indicate the coherence between the RAFs in the individual models. The lack of coherence between the eruptions indicates that the uncertainty

within a given model retrieval is greater than the uncertainty between models (i.e., the RAF of the Krakatoa eruption does not predict much of the RAF of another eruption for a given model).

We calculate a weighted mean volcanic forcing RAF of 0.80 using the total globally averaged aerosol optical depths from each eruption as weights. This is smaller than the 0.92 calculated from the GISS ModelE forcing from Hansen et al. (2005). Since volcanic eruptions are transient events, this does not affect future projections of the ERF. However, this does affect estimates of the cumulative energy that has come into the earth system. The CMIP5 multimodel mean volcanic forcing RAF of 0.8 leads to a difference in the integrated forcing between 1870 and 2012 of 8×10^{22} J just from the five largest volcanic events compared with the IPCC AR5. Our analysis suggests that about 5% more energy has
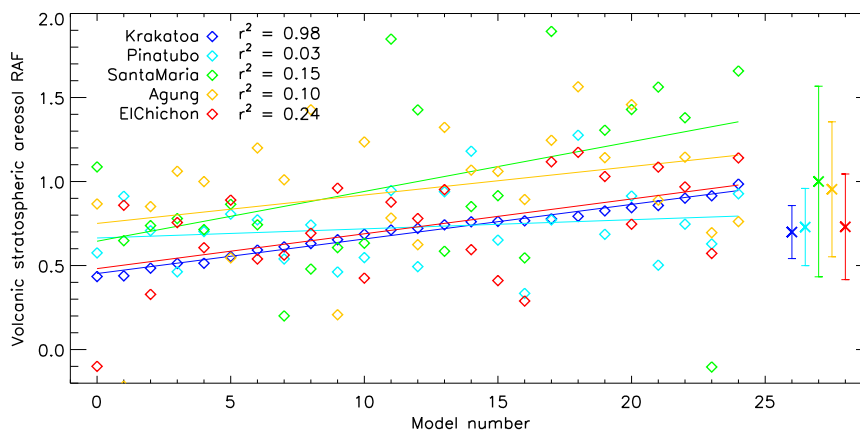

FIG. 9. Retrieved rapid adjustment factor (RAF) for each model and volcanic event sorted by the RAF of the Krakatoa eruption.

come into the earth system between 1870 and 2012 than suggested by the IPCC.

7. Conclusions

We have explored a new temporal kernel (TK) method to compute the effective radiative forcing (ERF) in transient simulations using an extension of the step-response method of Good et al. (2011, 2013). In addition, we have compared this method to an energy balance method (Forster and Taylor 2006) to compute ERF. Both methods allow the estimation of the ERF in transient simulations using only output from those simulations (along with control simulations needed to estimate needed inputs). In general these methods agree well, but they have different advantages and disadvantages.

The TK method takes the forward model of Good et al. (2011, 2013) that was developed to predict model response based from a forcing time series and inverts it to predict the forcing time series. In general, a step change experiment with output for the variable used in the inversion is needed to set up the model. The TK method is sensitive to the choice of variable used to derive the kernel and the noise in that variable in the transient simulation. We have identified the TOA radiative imbalance fit with a double exponential [see Eq. (6)] as a good choice to create the kernel function. Using the TOA imbalance, the TK method gives consistent results with the EB method and has some improved characteristics, including capturing the changing climate sensitivity in long simulations better. The largest downside to this method is that it is more susceptible to interannual variability, which can be amplified by the kernel compared with the EB method. However, this method of calculating ERF is likely more accurate for very long simulations of many centuries.

Regressions were originally used to calculate the forcing of step change experiments, and later the energy balance equation [see Eq. (1)] was directly applied to transient simulations using N and ΔT. This method assumes that the climate sensitivity parameter does not change with time or forcing agent. Generally, the climate feedback parameter is computed using a step change simulation from the same model using a regression of Eq. (1). Thus, the EB method is sensitive to the length of regression used to compute the climate feedback parameter.

The length of the regression used to calculate the inputs (α and F_0) for the two methods affects the energy budget estimated from the models. Between 1995 and 2015, using a climate feedback parameter estimated from the 20-yr regressions produces ERF estimates that average $0.2 \, \mathrm{W \, m^{-2}}$ higher than ERF estimated with

150-yr regression feedback parameters. Using 100-yr regressions retrieved a value for the climate feedback parameter that is closest to Murphy et al. (2009), which was derived from observations. The nonlinearity of the relationship between surface temperature change and TOA imbalance leads to biases in the retrieved forcing with the EB method. These biases are very model dependent. However, they usually grow with time, and ERF estimations for long simulations of many centuries can be greatly affected.

There are assumptions in these methods that lead to uncertainty and inaccuracy in ERF estimates. First, both methods assume that the climate responds uniformly to any global mean forcing, regardless of the type of forcing agent, state of the climate system, or location of the forcing. Forcing due to aerosols, land-use changes, solar radiation, etc. will have a different temperature and imbalance response than CO_2. CO_2 is the dominant forcing agent currently and in future scenarios; however, this assumption still leads to uncertainty in our ERF estimates. The TK method may be improved by combining step change experiments from other forcing agents or using efficacy estimates from Hansen et al. (2005). However, this is beyond the scope of this project, which is to introduce and test the TK method against the canonical energy balance method for estimating forcing in transient simulations. Second, the TK method assumes that the change in ERF in any given year can be combined linearly with previous and future ERF changes to estimate a response. Finally, both methods assume that the forcing response is the same for all future and past climate states as it is in the step change experiment. It is hard to quantify the effect of this assumption, but it is likely to be more important in higher RCP scenarios.

Other studies (such as Zhang and Huang 2014) have used radiative kernels to investigate the spatial pattern of forcing in CMIP5 models. The TK method described here could also be used on each grid cell to calculate a spatial map of the ERF; however, interannual variability could lead to large uncertainties. This may still be a valuable method for understanding regional forcing.

Both methods investigated have limitations for estimating effective radiative forcing in transient simulations, so we still need models to calculate their forcing online as much as possible. Much more work needs to be done to accurately derive and compare forcing across models. The Radiative Forcing Model Intercomparison Project (RFMIP), which is part of the suite of simulations for phase 6 of CMIP (CMIP6), will be an important next step toward this endeavor.

We also investigate the rapid adjustment factor (RAF) of volcanic aerosol with our two ERF estimation

methods. We calculate a CMIP5 multimodel mean RAF of 0.8 for the five largest volcanic eruptions since 1880, compared with Hansen et al. (2005) using GISS ModelE (0.9) and Meinshausen et al. (2011) (0.7). There is substantial variation in our calculated RAF between models and between volcanic events, which makes it difficult to draw robust conclusions. For instance, we do not detect a significant difference in RAF between the different eruptions, which might be expected based on their different latitudes and sizes. One important result of this analysis is that the CMIP5 multimodel mean RAF for volcanic aerosol forcing suggests less negative forcing during large eruptions than reported by the IPCC AR5. This has implications for the global energy budget. Using the RAF we compute for the CMIP5 multimodel mean, we estimate that approximately 5% more energy has come into the earth system since 1870 than reported by the IPCC AR5.

The temporal kernel method for calculating ERF is a new and useful tool that may demonstrate an improvement on current methods to calculate effective radiative forcing in transient climate simulations. This method becomes particularly useful for future predictions under high RCP scenarios where biases in the EB method may become large.

REFERENCES

Allan, R. P., C. Liu, N. G. Loeb, and M. D. Palmer, 2014: Changes in global net radiative imbalance 1985–2012. *Geophys. Res. Lett.*, **41**, 5588–5597, doi:10.1002/2014GL060962.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, doi:10.1029/2012GL051607.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter07_FINAL.pdf.]

Forster, P. M., and K. E. Taylor, 2006: Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *J. Climate*, **19**, 6181–6194, doi:10.1175/JCLI3974.1.

——, T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res. Atmos.*, **118**, 1139–1150, doi:10.1002/jgrd.50174.

Good, P., J. M. Gregory, and J. A. Lowe, 2011: A step-response simple climate model to reconstruct and interpret AOGCM projections. *Geophys. Res. Lett.*, **38**, L01703, doi:10.1029/2010GL045208.

——, ——, ——, and T. Andrews, 2013: Abrupt CO_2 experiments as tools for predicting and understanding CMIP5 representative concentration pathway projections. *Climate Dyn.*, **40**, 1041–1053, doi:10.1007/s00382-012-1410-4.

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, doi:10.1029/2003GL018747.

Gupta, A. S., N. C. Jourdain, J. N. Brown, and D. Monselesan, 2013: Climate drift in the CMIP5 models. *J. Climate*, **26**, 8597–8615, doi:10.1175/JCLI-D-12-00521.1.

Hansen, J., M. Sato, and R. Ruedy, 1997: Radiative forcing and climate response. *J. Geophys. Res.*, **102**, 6831–6864, doi:10.1029/96JD03436.

——, and Coauthors, 2002: Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.*, **107**, 4347, doi:10.1029/2001JD001143.

——, and Coauthors, 2005: Efficacy of climate forcings. *J. Geophys. Res.*, **110**, D18104, doi:10.1029/2005JD005776.

IPCC, 1990: *Climate Change: The IPCC Scientific Assessment*. Cambridge University Press, 410 pp.

——, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., doi:10.1017/CBO9781107415324.

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, doi:10.1007/s10584-011-0156-z.

Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, and T. Wong, 2009: An observationally based energy balance for the Earth since 1950. *J. Geophys. Res.*, **114**, D17107, doi:10.1029/2009JD012105.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, **98**, 22 987–22 994, doi:10.1029/93JD02553.

Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys. Res. Lett.*, **27**, 2685–2688, doi:10.1029/2000GL011373.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. *J. Climate*, **21**, 2269–2282, doi:10.1175/2007JCLI2044.1.

Shine, K. P., 2003: An alternative to radiative forcing for estimating the relative importance of climate change mechanisms. *Geophys. Res. Lett.*, **30**, 2047, doi:10.1029/2003GL018141.

Smith, D. M., R. P. Allan, and A. C. Coward, 2015: Earth's energy imbalance since 1960 in observations and CMIP5

models. *Geophys. Res. Lett.*, **42**, 1205–1213, doi:10.1002/2014GL062669.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Climate*, **21**, 3504–3520, doi:10.1175/2007JCLI2110.1.

Solomon, S., K. H. Rosenlof, R. W. Portmann, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner, 2010: Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science*, **327**, 1219–1223, doi:10.1126/science.1182488.

——, J. S. Daniel, R. R. Neely, J.-P. Vernier, E. G. Dutton, and L. W. Thomason, 2011: The persistently variable "background" stratospheric aerosol layer and global climate change. *Science*, **333**, 866–870, doi:10.1126/science.1206027.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Trenberth, K. E., and J. T. Fasullo, 2013: An apparent hiatus in global warming? *Earth's Future*, **1**, 19–32, doi:10.1002/2013EF000165.

Williams, K. D., W. J. Ingram, and J. M. Gregory, 2008: Time variation of effective climate sensitivity in GCMs. *J. Climate*, **21**, 5076–5090, doi:10.1175/2008JCLI2371.1.

Zhang, M., and Y. Huang, 2014: Radiative forcing of quadrupling CO_2. *J. Climate*, **27**, 2496–2508, doi:10.1175/JCLI-D-13-00535.1.